Machine Learning 1

Lecture 11.4 - Kernel Methods
Intermezzo: Constraint Optimization

Erik Bekkers

(Bishop E, 7.1)

Image credit: Kirillm | Getty Images

Slide credits: Rianne van den Berg

Intermezzo: Optimization with equality constraints

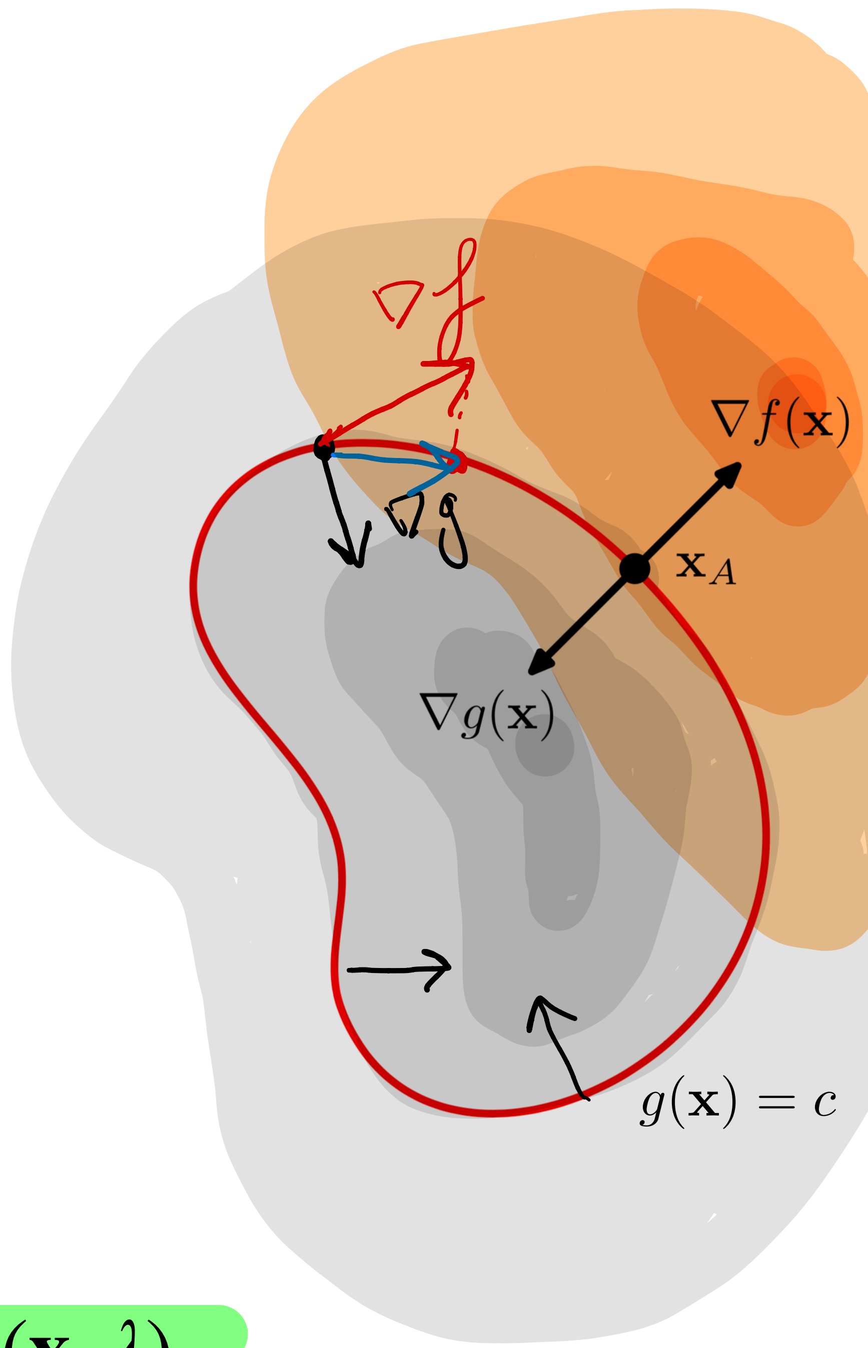
- Problem: Maximize $f(\mathbf{x})$ *subject to* $g(\mathbf{x}) = 0$
- Useful property: $\nabla g(\mathbf{x})$ is perpendicular to the constraint surface
- At constrained maximum, $\nabla f(\mathbf{x})$ must also be perpendicular to constraint surface
- Therefore: $\nabla f(\mathbf{x}) + \lambda \nabla g(\mathbf{x}) = 0$

  λ: Lagrange multiplier
- It is helpful to introduce a Lagrangian function:

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x})$$

- Solutions to original problem: stationary points of $L(\mathbf{x}, \lambda)$

$$\frac{\partial}{\partial \mathbf{x}} L(\mathbf{x}, \lambda) = 0, \qquad \frac{\partial}{\partial \lambda} L(\mathbf{x}, \lambda) = 0$$



Intermezzo: Optimization with inequality constraints

- Problem: Maximize $f(\mathbf{x})$ *subject to* $g(\mathbf{x}) \geq 0$ (1)

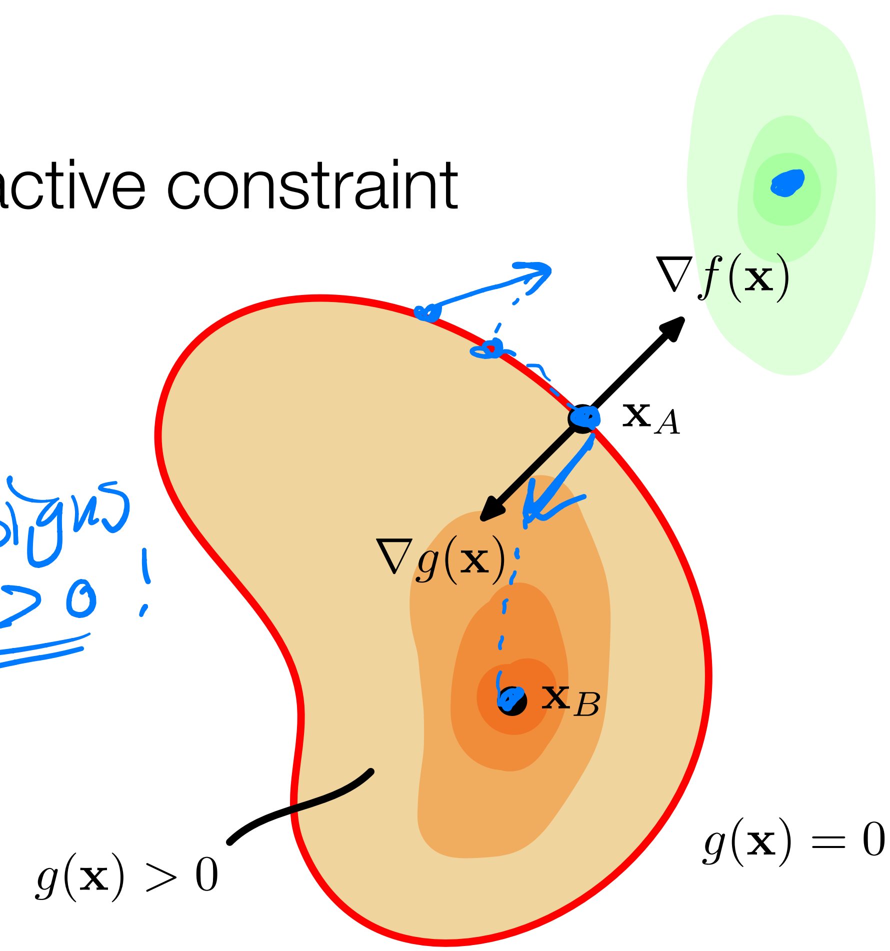
- Two kinds of solutions:
  - Stationary point lies in region $g(\mathbf{x}) \geq 0$: inactive constraint

$$\nabla f(\mathbf{x}) = 0, \qquad \mu = 0$$

  - Stationary point lies on boundary $g(\mathbf{x}) = 0$: active constraint

$$\nabla f(\mathbf{x}) = -\mu \nabla g(\mathbf{x}), \qquad \mu > 0$$

now the gradients must have opposite signs $\rightarrow \underline{\underline{\mu > 0}}$!



- **Primal Lagrangian**:

$$L(\mathbf{x}, \mu) = f(\mathbf{x}) + \mu g(\mathbf{x})$$

- Solution to (1):

$$\max_{\mathbf{x}} \min_{\mu} L(\mathbf{x}, \mu)$$ *subject to* **Karush-Kuhn-Tucker (KKT) conditions**

Complementary slackness

$$\mu \geq 0, \qquad g(\mathbf{x}) \geq 0, \qquad \mu g(\mathbf{x}) = 0$$

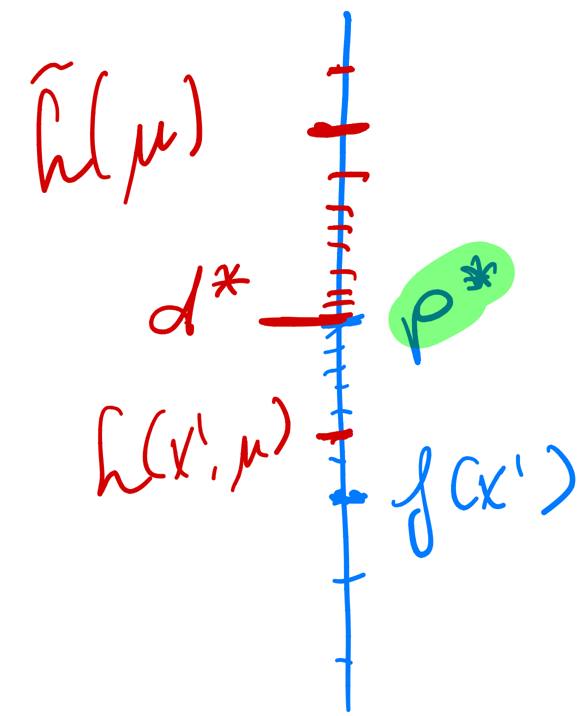
Intermezzo: Optimization with inequality constraints

- **Primal Problem**: $\max f(\mathbf{x})$ *subject to* $g(\mathbf{x}) \geq 0$ (1)
- Solution to (1): $\max_{\mathbf{x}} \min_{\mu} L(\mathbf{x}, \mu)$ *subject to*

KKT conditions
$\mu \geq 0, \;\; g(\mathbf{x}) \geq 0, \; \mu g(\mathbf{x}) = 0$

- **Dual Lagrangian** (Optimize w.r.t. primal variables $\mathbf{x}$ for fixed dual variables μ)

$$\tilde{L}(\mu) = \max_{\mathbf{x}} L(\mathbf{x}, \mu) \qquad \text{with} \qquad L(\mathbf{x}, \mu) = f(\mathbf{x}) + \mu g(\mathbf{x})$$



- Obtain dual Langrangian analytically:
  - Use stationarity condition $\nabla_{\mathbf{x}} L = 0$ to eliminate $\mathbf{x}$ from L
  - This gives $\tilde{L}$ which now only depends on μ
  - This is an upper bound for (1) as function of μ
- **Duality gap**: $d^* - p^*$
  - For every $\mathbf{x}'$ satisfying $g(\mathbf{x}') \geq 0$ we have $f(\mathbf{x}') \leq L(\mathbf{x}', \mu) \leq \tilde{L}(\mu)$
  - It follows (weak duality): $\mathbf{p}^* = \max_{\mathbf{x},\, g(\mathbf{x}) \geq 0} f(\mathbf{x}) \leq \min_{\mu} \tilde{L}(\mu) = \mathbf{d}^*$

Intermezzo: Optimization with inequality constraints

- **Primal Problem**: $\max f(\mathbf{x})$ *subject to* $g(\mathbf{x}) \geq 0$ (1)
- Solution to (1): $\max_{\mathbf{x}} \min_{\mu} L(\mathbf{x}, \mu)$ *subject to*

KKT conditions
$\mu \geq 0, \;\; g(\mathbf{x}) \geq 0, \; \mu g(\mathbf{x}) = 0$

- For (almost all) convex problems
  - Strong duality: $\mathbf{p}^* = \mathbf{d}^*$
  - So if we have solved the dual problem, we have solved the primal problem!
- **Dual problem** (find the lowest upper bound):

$$\min_{\mu} \tilde{L}(\mu) \textit{ subject to } \mu \geq 0$$

- Recipe:
  - Define Lagrangian $L(\mathbf{x}, \mu) = f(x) + \mu g(\mathbf{x})$
  - Compute dual Lagrangian $\tilde{L}(\mu)$
  - Solve dual problem: $\mu^* = \arg\min_{\mu} \tilde{L}(\mu)$ *subject to* $\mu \geq 0$
  - Maximize primal Lagrangian: $\mathbf{x}^* = \arg\max_{\mathbf{x}} L(\mathbf{x}, \mu^*)$